

Is the “Big Bang” in Animal Evolution Real?

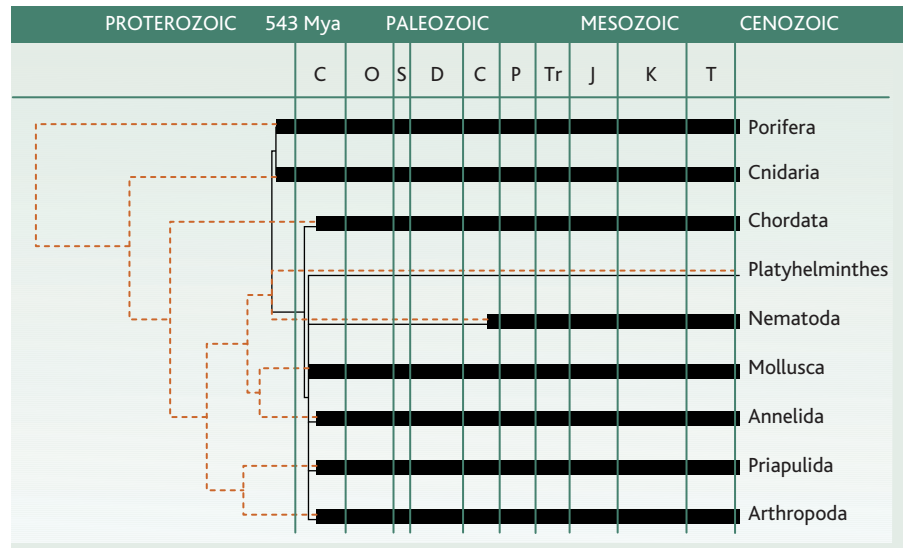
Lars S. Jermiin, Leon Poladian, Michael A. Charleston

The origin and evolution of animals have remained hotly debated issues ever since Darwin drew attention to the relative paucity of fossils from the Precambrian, which ended 543 million years ago (Mya) (1). On the one hand, a growing collection of exquisitely preserved fossils of soft-bodied animals from the Cambrian has highlighted the existence of Cambrian representatives of most of the living animal phyla (2). This has given rise to the “Cambrian Explosion” hypothesis (3) that most animal phyla arose ~543 Mya within a short period (see the figure). On the other hand, studies of increasingly large molecular data sets suggest that many of the same phyla arose in the Precambrian (4–6), leading to considerable and occasionally intemperate debate on the timing of divergence of animal phyla. The article by Rokas *et al.* (7) on page 1933 in this issue presents a new aspect to this controversy.

Motivated by the desire to resolve the early evolution of animals, while accounting for insufficient data from some taxa, Rokas *et al.* used an alignment of 12,060 amino acids, encoded by 50 genes, to infer the phylogeny of 16 representatives from nine animal phyla. Eleven of these were from Porifera, Cnidaria, Platyhelminthes, Mollusca, Annelida, and Priapulida. These six phyla have not been adequately represented in recent molecular studies of early animal evolution (4–6). Given the aligned data, Rokas *et al.* inferred a phylogeny with several distinct speciation events that were consistently supported by the data, but with a conspicuous polytomy (multiple, concurrent divergence events) involving the proto-stome phyla, and another involving the bilaterate, cnidarian, and poriferan lineages. Recognizing that a polytomy may be due to rapid speciation, or to poor or insufficient data, Rokas *et al.* then explored whether

convergence, “rogue” sequences, compositional heterogeneity, missing or inadequate data, or mutational saturation could have affected the phylogenetic estimate, and found that none of these were likely. In the absence of other explanations, they concluded that their data support rapid speciation during the early evolution of animals.

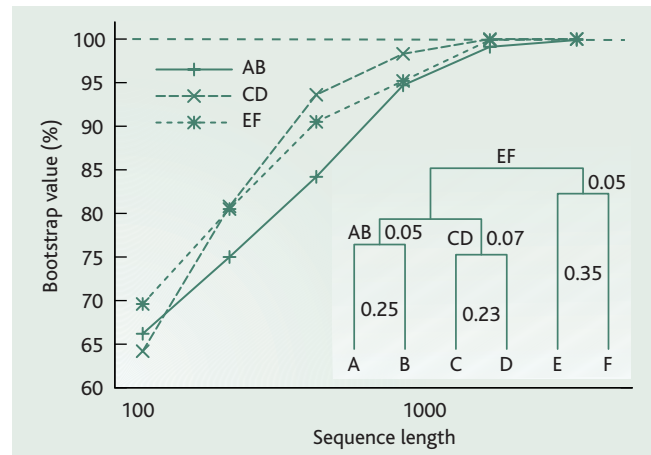
This conclusion is significant because it is consistent with the Cambrian Explosion hypothesis. Whereas previous molecular studies have concluded that the divergence of animal phyla occurred gradually over a period stretching hundreds of millions of years into the Precambrian (4–6), the present study suggests that two episodes of multiple divergence events occurred, each over just a few million years. This conclusion is consistent with a phylogeny based on fossil data (see the figure), thus reconciling a key difference in opinion between evolutionary biologists and paleontologists. However, Rokas *et al.* did not date these episodes, so corroboration of the Cambrian Explosion hypothesis is conditional on future estimates of the



The debate on early evolution of animals. (Top)

The fossil record and evolution of 9 of the 35 currently recognized metazoan phyla suggest that most animal phyla diverged/arose at the beginning of the Cambrian (C) period. The thick lines represent the known ranges of fossils from their first appearance in the fossil record (16). Thin lines represent the inferred metazoan phylogeny based on fossil data. Dashed lines represent an amalgam of

three conservative estimates of the inferred metazoan phylogeny (4–6). (Bottom) Demonstration of the increase in bootstrap values when longer sequences are used. The phylogenetic tree shown (inset) is that used to generate six nucleotide sequences labeled A to F, each of length 100; these were duplicated to a length of 200, 400, 800, 1600, and 3200 bases. The internal branches are labeled AB, CD, and EF; the numbers at each branch are proportional to time. The data were generated under a Markov model that allows two classes of substitutions, using Seq-Gen (17); bootstrap values were estimated with PAUP* (18).



L. S. Jermiin is at the School of Biological Sciences, L. Poladian is at the School of Mathematics and Statistics, and M. A. Charleston is at the School of Information Technology, University of Sydney, NSW 2006, Australia. E-mail: lars.jermiin@usyd.edu.au All authors are members of the Sydney University Biological Informatics and Technology Centre (SUBIT), University of Sydney, NSW 2006, Australia.

divergence dates not falling well within the Precambrian.

Is the conclusion drawn by Rokas *et al.* sound? At first glance, it appears so, but can their conclusion stand up to closer scrutiny? Accepting that the genes analyzed by the authors evolved without gene duplication and that the amino acids are aligned correctly, most phylogenetic methods assume that the evolutionary dynamics of the 12,060 amino acid sites are independently and identically distributed, and that they evolved under the same stationary, reversible, and homogeneous conditions (8). The assumptions arise from the need to render phylogenetic methods tractable and easy to use, and they are unlikely to be realistic. To account for the observation that the sites in a gene may evolve at different rates, some phylogenetic methods are able to model rate heterogeneity across sites using a Γ distribution (9). Rokas *et al.* used this approach for the whole alignment but did not consider that different parts of the alignment may require different Γ distributions. Nor did they consider that some sites may vary nonindependently (10) and that the distribution of variable sites may vary across lineages and through time, an issue that is notoriously difficult to resolve (11). A logical extension to the work would be to partition the alignment and estimate the evolutionary rates for different genes separately.

Violation of the assumed stationary, reversible, and homogeneous conditions may lead to compositional differences in the aligned amino acid sequences and hence to errors in phylogenetic estimates (12). Rokas *et al.* recognized this potential source of error but used a test that is known to be flawed, even though better tests are known (13). Furthermore, they chose a phylogenetic method that, while it accounts for compositional variation in the sequence alignment, is unsuitable: It assumes that the sites are independently and identically distributed, which they have already shown not to be the case. Moreover, they used a single Markov (probabilistic) model to analyze the alignment of amino acids, in effect using a "one size fits all" approach, where it would have been better to use several Markov models to capture gene-specific differences in the evolutionary processes (14).

Rokas *et al.* used nonparametric bootstrap and posterior probabilities to gauge support for the pattern and order of speciation events (branches in their phylogenetic tree). The former is widely recognized to be statistically unwise. Bootstrap values are estimates of the expected frequency with which speciation events (internal branches) occur in the optimal tree, using data constructed from the original alignment by sam-

pling sites with replacement (15). It is not a measure of accuracy or confidence, but of data consistency. Further, the increase in bootstrap value when more genes are included may be misleading, because longer sequences naturally tend to have higher bootstrap values (see the figure). The posterior probabilities of speciation events being correctly identified are also prone to error when the phylogenetic assumptions are violated in the sense described above.

In light of these concerns, are the conclusions of Rokas *et al.* justified? Should we ignore their study? Most certainly not, because they have produced a wealth of data and have shown that it might just be possible that the fossil record can be reconciled with molecular data. This, in itself, should be cause for celebration and an incentive to acquire sequence data from the remaining 26 animal phyla. Likewise, it should encourage development of methods that assess when data violate phylogenetic assumptions, and that cope with such data. To achieve these goals, we need to know more about the structure and function of gene products before we can develop models that appropriately address the early evolution of animals.

PHYSIOLOGY

The Tick-Tock of Aging?

Adam Antebi

The relationship between organismal development and aging has long been a matter of intense debate. It seems natural to posit that developmental timing mechanisms that culminate in reproductive maturity continue to affect post-reproductive biology, with consequences for total organism life span. On the other hand, evolutionary theories of aging discount regulated aging per se, because the force of selection declines with age and drops precipitously after reproductive potential ends. In other words, a program that actively ages the organism is unlikely to be selected for in evolution. Instead, aging is thought to entail the passive stochastic accumulation of damage to molecules, cells, and organs, leading to loss of fertility and organismal demise. Therefore, the notion that regulated intrinsic biological timers control aging seems superficially untenable.

Just this possibility, however, has been raised by Boehm and Slack on page 1954 of this issue (1). They have found that compo-

References

1. C. Darwin, *The Origin of Species by Means of Natural Selection*. (John Murray, London, ed. 6, 1888), p. 313.
2. D. E. G. Briggs, R. A. Fortey, *Paleobiology* **32**, 594 (2005).
3. S. J. Gould, *Wonderful Life: The Burgess Shale and the Nature of History* (Hutchinson Radius, London, 1989).
4. E. J. Douzery, E. A. Snell, E. Bapteste, F. Delsuc, H. Philippe, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15386 (2004).
5. K. J. Peterson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6536 (2004).
6. K. J. Peterson, N. J. Butterfield, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9547 (2005).
7. A. Rokas, D. Krüger, S. B. Carroll, *Science* **310**, 1933 (2005).
8. V. Jayaswal, L. S. Jermini, J. Robinson, *Evol. Bioinformatics Online* **1**, 62 (2005).
9. Z. Yang, *Trends Ecol. Evol.* **11**, 367 (1996).
10. D. Penny, B. J. McComish, M. A. Charleston, M. D. Hendy, *J. Mol. Evol.* **53**, 711 (2001).
11. P. J. Lockhart *et al.*, *Mol. Biol. Evol.* **23**, 40 (2006).
12. S. Y. W. Ho, L. S. Jermini, *Syst. Biol.* **53**, 623 (2004).
13. L. S. Jermini, S. Y. W. Ho, F. Ababneh, J. Robinson, A. D. W. Larkum, *Syst. Biol.* **53**, 638 (2004).
14. M. Pagel, A. Meade, *Syst. Biol.* **53**, 571 (2004).
15. J. Felsenstein, *Evolution* **39**, 783 (1985).
16. J. W. Valentine, D. Jablonski, D. H. Erwin, *Development* **126**, 851 (1999).
17. A. Rambaut, N. C. Grassly, *Comp. Appl. Biosci.* **13**, 235 (1997).
18. D. L. Swofford, *Phylogenetic Analysis Using Parsimony (*and Other Methods)*, 4.0 Beta (Sinauer, Sunderland, MA, 2002).

10.1126/science.1122440

nents of a nematode's (*Caenorhabditis elegans*) heterochronic circuit—namely the *lin-4* microRNA and its target, the nuclear protein encoded by *lin-14*—not only perturb developmental timing but also influence organismal life span. They do so by regulating insulin/IGF-1 (insulin-like growth factor-1) signaling, a cellular regulatory pathway whose modest decrease in activity leads to increased longevity across taxa (2).

Just as each cell in a developing organism has a positional identity that is determined by gradients of morphogens and hierarchies of transcription factor activity, cells also have a temporal identity dictated by regulatory signaling cascades. Pioneering work in *C. elegans* led to the discovery of the heterochronic loci (3), which constitute a regulatory circuit that confers temporal identity to the various tissues. These genes determine cellular programs of division, migration, and differentiation that are appropriate for a specific developmental stage. In addition, the interactions among these heterochronic loci ensure the proper succession of larval temporal fates. Normally, *C. elegans* develops to adulthood through four larval stages, L1 to L4. Worms with mutations in the heterochronic loci inappropriately express cel-

The author is at the Huffington Center on Aging, Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA. E-mail: aantebi@bcm.tmc.edu